

CRAWLING WEBSITE E-GOVERNMENT PEMERINTAH DAERAH JAWA TENGAH BERBASIS ONTOLOGY

BIMA JATI WIJAYA

*Program Studi Teknik Informatika - S1, Fakultas Ilmu
Komputer, Universitas Dian Nuswantoro Semarang*

URL : <http://dinus.ac.id/>

Email : 111201105929@mhs.dinus.ac.id

ABSTRAK

Agregasi informasi pada situs e-gov Indonesia sangat diperlukan untuk memelihara konten portal web dan memenuhi kebutuhan informasi yang kompleks serta banyaknya jumlah data saat ini sulit untuk mendapatkan informasi yang relevan. Oleh karena itu harus dilakukan crawl pada website pemerintah dengan membuat program crawl. Akan tetapi crawl pada umumnya mendownload seluruh url tanpa melihat content didalamnya sedangkan crawl terfokus memiliki keterbatasan tidak memiliki cakupan yang luas dalam suatu konsep. Oleh karena itu dibutuhkan crawling berbasis ontologi untuk menyelesaikan masalah tersebut dengan disertai pengecekan relevansi menggunakan tfidf dan cosine similarity. Dengan pendekatan ontologi serta pengecekan relevansi tfidf dan cosine similarity diharapkan menghasilkan crawler yang effesien serta mempermudah dalam penggunaan hasil crawling karena sudah terklasifikasi.

Kata Kunci : Web crawl, crawler berdasarkan ontology, tf-idf

WEBSITE CRAWLING ON E-GOVERNMENT OF CENTRAL JAVA GOVERNMENT BASED ON ONTOLOGY

BIMA JATI WIJAYA

*Program Studi Teknik Informatika - S1, Fakultas Ilmu
Komputer, Universitas Dian Nuswantoro Semarang*

URL : <http://dinus.ac.id/>

Email : 111201105929@mhs.dinus.ac.id

ABSTRACT

Aggregation of information on e-government Indonesian very required to maintain the web portal content and meet complex information needs and currently difficult to obtain relevant information from the large amount of data. Therefore it have to do crawl on the e-government website with create web crawl. However generally crawl download entire URL without seeing the content. While focused crawl has a limitation in the concept coverage. The issue became the basis of research to make web crawler based on ontology to solve the issue which combine with relevant check content use tf-idf and cosine similarity method. The existing approach on web crawler based ontology use tfidf and cosine similarity method is expected to solve the issue and it will becomes efficient web crawler which make easier to reprocess from result of crawling because it was classified.

Keyword : Web crawl, crawler based ontology, tf-idf